

Concordance of Actigraphic and Ecological Momentary Assessed Sleep

by

Jocelyn Nicole Mineo

BS, Bowling Green State University, 2017

Submitted to the Graduate Faculty of
the Department of Biostatistics
the Graduate School of Public Health in partial fulfillment
of the requirements for the degree of
Master of Science

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH
GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Jocelyn Nicole Mineo

It was defended on

April 12, 2019

and approved by

Thesis Advisor: Ada Youk, PhD, Associate Professor, Department of Biostatistics, Department of Epidemiology, Department of Clinical and Translational Science, Graduate School of Public Health, University of Pittsburgh

Christopher Kline, PhD, Assistant Professor, Department of Health and Physical Activity, School of Education, University of Pittsburgh

Robert Krafty, PhD, Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Copyright © by Jocelyn Nicole Mineo

2019

Concordance of Actigraphic and Ecological Momentary Assessed Sleep

Jocelyn Nicole Mineo, MS

University of Pittsburgh, 2019

Abstract

There are many studies that examine different forms of sleep measurement in order to understand sleep patterns within different populations, including those with mental health issues, sleep disorders, chronic pain, and adolescents. These studies often compare subjective and objective methods of sleep measurement to observe whether or not they are in agreement. In this thesis, we examine concordance between a subjective sleep measurement, in the form of a questionnaire, and an objective sleep measurement, in the form of an actigraphic sleep watch. Four subjective questions were paired with four actigraphic sleep measurements for analysis. Continuous actigraphic measurements were categorized into quartiles or deciles to compare with their subjective categorical counterparts. Concordance was tested using three-level nested longitudinal models and Bland-Altman plots. We also analyze these variables to check for a directional influence over time using cross-lag regression methods. The intraclass correlations from the longitudinal analysis and the Bland Altman plots both revealed that the two methods of sleep measurement are not concordant. The cross-lag regression analysis indicated that there is a statistically significant directional influence for the participants' number of awakenings and quality of sleep. Our findings supplement the findings of prior research about the lack of concordance of sleep measurement, while also adding to this work by considering directional influence.

Public Health Significance: Sleep habits can impact a person's immune system, metabolism, autonomic nervous system activity, and mental health. Understanding if different methods of sleep measurement are concordant is necessary for future investigators to assess the accuracy of each measurement and choose which measurement is the most appropriate for their research. The cross-lag regression methods used in this analysis could be utilized by future researchers to understand the directional influence of other sleep variables over time.

Table of Contents

Preface.....	x
1.0 Introduction.....	1
1.1 Background.....	1
1.2 Literature Review	2
2.0 Methods.....	8
2.1 Study Population	8
2.2 Primary Variables of Interest.....	9
2.3 Statistical Analyses	10
2.3.1 Descriptive Statistics	10
2.3.2 Testing Concordance	10
2.3.3 Cross-lag Regression	13
3.0 Results	15
3.1 Descriptive Statistics	15
3.2 Testing Concordance.....	17
3.3 Cross-lag Regression	25
4.0 Discussion.....	27
Appendix: Statistical Code.....	29
A.1 Additional Figures	29
A.1.1 Histograms of the Objective Sleep Measurement Variables	29
A.2 Stata Example Code	30
A.2.1 Stata example code for descriptive analysis.....	30

A.2.2 Stata example code for three-level random intercept model.....	30
A.2.3 Stata example code for Bland-Altman plots²².....	30
A.2.4 Stata example code for cross-lag regression analysis.....	32
Bibliography	34

List of Tables

Table 1. Descriptive statistics for demographic and sleep variables.	16
Table 2. Categorized variable counts.....	17
Table 3. Intraclass correlations for each model.	18
Table 4. Summary of cross-lag regression for each model.....	26

List of Figures

Figure 1. A visual representation of the cross-lag regression model.	6
Figure 2. Three-level random intercept model.	11
Figure 3. Residual diagnostics of three-level random intercept model using Empirical Bayes prediction.	22
Figure 4. Bland-Altman plots that show the average between methods versus the difference between methods.	24
Figure 5. Distributions of the objective sleep measurements.	29

Preface

First of all, I would like to thank my academic advisor and thesis supervisor, Dr. Ada Youk for answering countless emails, sitting through all of my long meetings, and being an overall great source of support and guidance during my time at the University of Pittsburgh.

I also want to thank Dr. Christopher Kline and Dr. Robert Krafty for agreeing to be a part of my thesis committee in the first place and working with me throughout this project. I want to thank the entire faculty in the Biostatistics department at the University of Pittsburgh. I have gained so much valuable experience during my time here that would not have been possible without them.

Thank you to Dr. Lora Burke and Dr. Stephen Rathbun for providing me with and helping me to understand their data. I appreciate both of you for making sure I was equipped with what I needed in a timely manner from the start of my work on this project.

I would also like to thank the extremely supportive friends I have made during my time here, Avantika and Michelle. Thank you for always being there for me, whether it was listening to my thoughts or distracting me from them, I could not have done it without both of you.

Of course, last, but not least, I want to thank those who are the most important to me. To my parents, I would not be who I am or where I am today without your constant encouragement. To my brother, thank you for urging me to change my major during my time as an undergraduate all those years ago. If not for your advice, I could not have made it here. To Zachary, thank you for your consistent support, love, and for always believing I could do anything, even when I did not.

1.0 Introduction

1.1 Background

Understanding sleeping habits is essential to understanding other aspects of health. Recent literature has linked sleep to the formation and maintenance of immunological memory, metabolism, autonomic nervous system activity, and mental health.¹ Understanding sleeping patterns is also important to populations that struggle to get enough sleep, such as those with insomnia or chronic pain. Several studies have utilized both subjective and objective measurements of sleep. The most common type of subjective sleep measurement is a sleep diary or questionnaire filled out soon after waking up. One of the main shortcomings of subjective methods of sleep measurement is that an individual may not be able to accurately assess their own sleep. Some forms of objective sleep measurement include actigraphy, also known as a wrist-worn accelerometer, and polysomnography (PSG). While PSG is considered to be the gold-standard for measuring sleep objectively, actigraphy is often preferred because it is a less invasive and expensive way to obtain an objective measure of sleep.² Due to the numerous health implications of getting enough sleep, it is important to understand the differences between these forms of sleep measurement, the drawbacks of each one, and which is most reliable and valid.

Recent literature has reported analyses to study the differences between these two methods of sleep measurement. These studies examine the differences between these types of measurements in the total amount of sleep recorded, number of times a person wakes up while sleeping, total amount of time awake throughout the night, and more. The studies have found that subjective methods tend to underestimate wake after sleep onset in comparison to PSG. Another consistent

finding among these studies is that subjective measures tend to have much higher estimates of sleep onset latency than objective measures.³ There has also been research that examines the differences between the two objective sleep measurements. Researchers have shown that actigraphy tends to overestimate the amount of sleep a person gets when compared to PSG because actigraphy does not do as well at detecting wakefulness.⁴

In this thesis, a subjective and an objective method of sleep measurement are compared in order to further understand which method is more reliable and valid. The subjective method of measuring sleep used is a questionnaire that was collected in the morning by ecological momentary assessment (EMA), which is the repeated assessment of the participants' behaviors and experiences in real time in their natural environments. The main strength of EMA is the minimization of recall bias from participants.⁵ The objective method of measuring sleep utilizes actigraphic data on the duration of sleep and the number of awakenings. Data were collected for a span of seven days, at six months and again at twelve months. Concordance between the questionnaire and actigraphic data will be assessed at each time point by utilizing a repeated measures ANOVA and examining intraclass correlations. In order to visually assess concordance, Bland-Altman plots will be created. In addition, this thesis will attempt to assess whether changes in subjective sleep data precede changes in the actigraphy data utilizing cross lag regression methods.

1.2 Literature Review

The EMPOWER study collected data on a multitude of topics, including physical activity and eating habits, weight loss, motivation levels, goals of the participants, and, the primary focus of

this thesis, sleeping habits. The diverse data collected has allowed investigators to conduct secondary analyses in several different fields of study. One study utilized the data on participants' motivation, goals, and weight loss to evaluate if self-weighing led to weight loss and step goals.⁶ Another study used variables from this dataset to understand if there is a bidirectional relationship between obstructive sleep apnea and changes in weight.⁷ Nursing researchers analyzed behavior and outcome changes in participants using data on their caloric intake and weight using group-based trajectory modeling.⁸ Neighborhood environments of the participants were assessed in another study to examine if there is an association between changes in weight and neighborhood factors, such as racial composition, neighborhood income, and density of food retail stores.⁹ It is evident that the EMPOWER study has made a large contribution to many fields of research, but what has not been assessed yet is the concordance between the different methods of sleep measurement that were used in this study.

Concordance, or agreement, between sleep measurements can be estimated using several different methods. Previous studies that have been done to measure concordance between sleep measurements often analyze the most popular objective methods, actigraphy and PSG, or they compare both objective methods of sleep measurement to a subjective method of sleep measurement. Some of the different subjective methods that are often included in these studies include questionnaires or sleep diaries that are recorded directly by the participant or researchers. When PSG is included in a study, researchers often calculate specificity, sensitivity, negative predictive values, and positive predictive values. These calculations provide an understanding of if actigraphy correctly classifies sleep and wake times, assuming PSG is correct because it is considered the gold standard. Prior studies trying to understand concordance between sleep

measurements have used statistical methods such as repeated measures ANOVAs, Bland-Altman plots, paired t-tests, Wilcoxon sign-rank tests, and survival agreement plots.

All of these methods adequately evaluate concordance, but in this thesis we chose a three-level random intercept model because it allows us to assess intraclass correlations for each level of the model that are calculated using the variance components from each level. The intraclass correlations tell us if the two methods of sleep measurement are concordant. For our analysis, level-one is the timepoint at which sleep was measured, level-two is the method of sleep measurement, and level-three is the subjects of the study (specific details of our models can be found in Section 2.3.2). Assumptions of this model include: error term components are normally distributed with a mean of zero, all three of the error components are uncorrelated across subjects, level-two random intercepts and level-one residuals are uncorrelated across methods of sleep measurement, level-one residuals are uncorrelated across timepoints, covariance of any two observations within the same subject using different methods, covariance of any two observations within the same subject using different methods, and observations from different subjects are uncorrelated. There are different methods to calculate residuals for this model, including the ordinary least squares method and the Empirical Bayes method. We use the Empirical Bayes method for the predictions of the residuals to assess model fit because they are more precise.¹⁰ We also chose Bland-Altman plots to visually assess concordance based on the width of the limits of agreement, the number of points outside of these limits, and if there is clustering within these limits. The graph plots the differences between the two methods of sleep measurement versus the average of the two methods, one of the methods, the sample rank, or the geometric mean of both of the methods. We chose to plot the differences between the two methods against the averages of the two methods. Our data are not ranked, so we cannot plot the differences against the sample

rank. Neither of the sleep measurement methods in this thesis are considered the gold standard, so it would not make sense to plot the differences against one of the methods. Two of our Bland-Altman plots included variables with a limited range. In order to supplement these plots, we included bar charts that will add to the understanding of the differences between the two methods of sleep measurements.¹¹

Another way to further understand the relationship between these methods of sleep measurement is to check for a time effect. Cross-lag regression analysis, which is also known as temporal causality, cross lag panel analysis, and Granger causality, is used to evaluate if there is causal predominance, which is when one variable has an effect on another without a reciprocal effect in return¹². In the structural model below, β_1 and β_2 represent the cross-lag effects, β_3 and β_4 represent auto-regressive paths, and r_{xy} represents the correlations between the two methods at the same time point.¹² In this thesis, cross-lag effects show the relationship between the two different forms of sleep measurement across two timepoints, and autoregressive paths show the relationships between the same form of sleep measurement over two timepoints. A diagram of the model used in our analysis can be seen below in Figure 1.

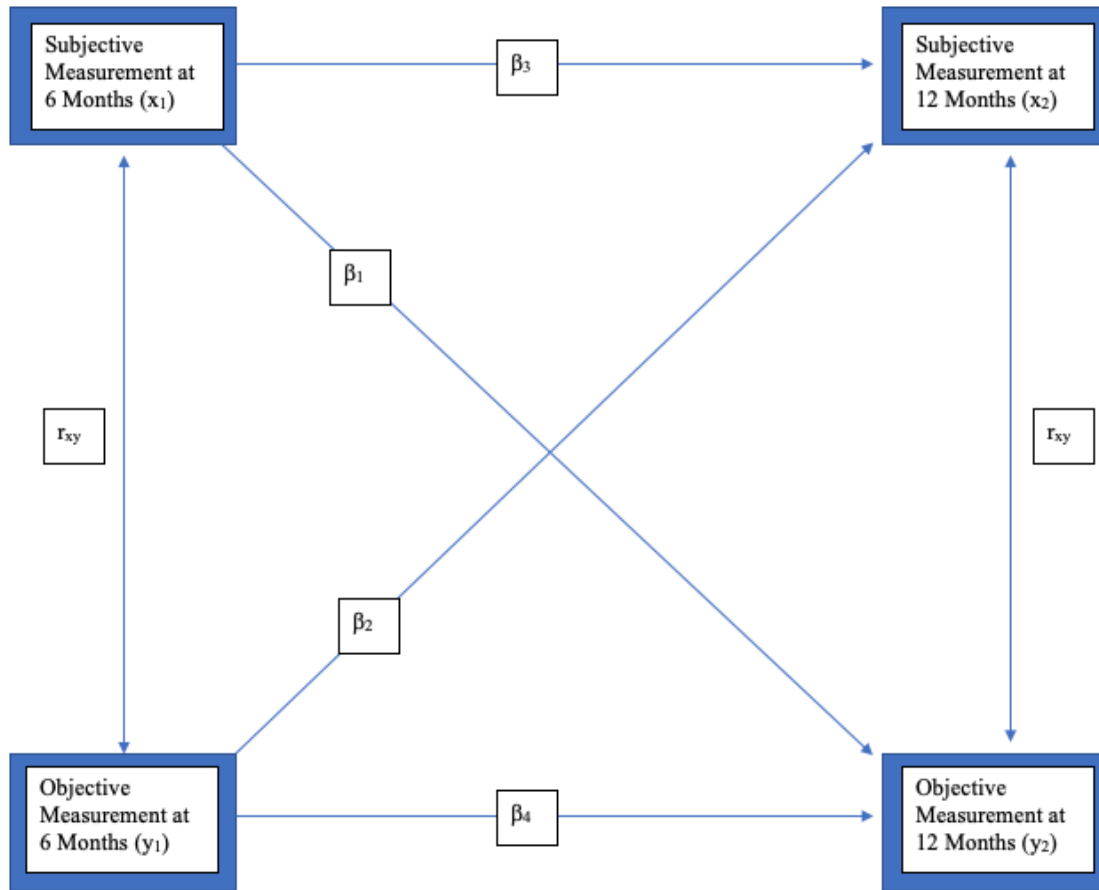


Figure 1. A visual representation of the cross-lag regression model.

These types of methods are often used in neuroscience and economics to understand causal influences. This model assumes synchronicity, stationarity, variables are measured without error, X_1 occurs before X_2 , no confounders exist, standardized variables, and stability. The synchronicity assumption is assuming that each measurement of sleep happened simultaneously at the two timepoints.¹² Stationarity and stability both relate to lack of change in variables over time. Stationarity refers to the causal processes, in terms of strength and direction, remaining the same at each timepoint. For perfect stationarity, the correlations between the two different methods of sleep measurement at the same time point (r_{xy}) would not change between timepoints. Stability refers to the extent to which the empirical value of a

variable stays the same at each time point.¹³ Autoregressive effects (β_3 and β_4) describe the stability of individual differences at each timepoint.¹² Standardizing variables is typically done for interpretability, but it is recognized that this is not always appropriate. In this thesis, we manipulate the continuous actigraphic variables by categorizing them based on quartiles or deciles, so they are all measured on the same scale. Three popular goodness of fit measurements are (1) root mean squared error of approximation, (2) the Tucker Lewis Index, and (3) the comparative fit index. The equations to calculate each of these measurements are shown below.¹⁴ A model that fits well will have a root mean squared error of approximation that is 0.05 or less, and a model is considered to have an exceptionally poor fit if the root mean squared error of approximation is greater than 0.10. The Tucker Lewis Index and the comparative fit index will be close to one if the model fits well, but either measure being below 0.90 indicates a poor model fit.¹⁴

$$(1) \text{ RMSEA} = \sqrt{\frac{\frac{\chi^2_1}{df_1} - 1}{N - 1}}$$

$$(2) \text{ TLI} = \frac{\frac{\chi^2_0}{df_0} - \frac{\chi^2_1}{df_1}}{\frac{\chi^2_0}{df_0} - 1}$$

$$(3) \text{ CFI} = \frac{(\chi^2_0 - df_0) - (\chi^2_1 - df_1)}{(\chi^2_0 - df_0)}$$

Goodness of fit results can be a tool to show not only how well a model fits, but also if model saturation has occurred. Model saturation means the model is a perfect fit, so the data will fit perfectly at any point. This negates the need for goodness of fit measurements altogether. A saturated model is not a concern if the model fits with the theoretical perspective of the analysis and is interpretable, and all of the models in this analysis are interpretable.

2.0 Methods

2.1 Study Population

Data are from an Ecological Momentary Assessment (EMA) study that was led by researchers at the University of Pittsburgh School of Nursing. The EMA analysis was a secondary analysis that utilized data from the EMPOWER study, which was a 12-month prospective observational study that aimed to understand triggers of lapses and relapses after intentional weight loss through EMA. Data collection included event contingent, signal contingent, and time contingent data. Event contingent data were collected through a smartphone EMA application when a predefined event occurred.¹⁵ Signal contingent data were collected at random time signified by a beep from their device, so researchers could understand how participants were feeling, their moods, and their environment. These assessments were given based on a known probability-based sampling design in order to be sure these data were collected randomly.¹⁵ Time contingent data were collected by giving individuals a fixed time and prompt to help researchers understand sleep patterns, how the individual was feeling that morning, and how they had felt about their daily goals.¹⁵

The study sample consisted of 150 individuals (136 females, 14 males), and one individual who withdrew from the study immediately after baseline. 121 of the individuals were white, and the average age was 51.09. The retention rate was 87.4%, so the study ended with 132 participants who remained in the study until completion. Reasons for individuals leaving the study included pregnancy, development of diabetes, personal decisions, and loss to follow-up.¹⁵

2.2 Primary Variables of Interest

As part of the collection of the time contingent data for the EMPOWER study, there were beginning of the day questions, as well as end of the day questions. This thesis utilizes only the beginning of the day questions related to the participants' sleep from the prior night. These questions were answered every morning for the full twelve months of the study. These questions included: (1) Did you have trouble falling asleep last night? (0 = no trouble at all, 10 = a lot of trouble) (2) How many hours of sleep did you get? (3) Number of awakenings? (4) Rate how well you slept last night? (0 = poor, 10 = excellent) (5) Do you feel tired (y/n)? This questionnaire was self-reported data, and it will serve as the main subjective sleep measurement in this study.

Actigraphic sleep watches were worn by participants for a period of seven days at six and twelve months to capture sleep duration and awakenings. In order to accurately assess concordance, actigraphy variables were chosen that align closely with the subjective questionnaire. These variables include (1) sleep onset latency, (2) total sleep time, (3) number of wake bouts, and (4) sleep efficiency. These actigraphic measurements will serve as the main objective sleep measurements for this thesis. For comparability to the subjective questions that were collected as a scaled response between zero and ten, the actigraphic variable will be divided into categories based on percentiles. The actigraphic variable sleep onset latency was heavily right-skewed, which required us to categorize this variable based on quartiles, and in order to analyze this variable with the first subjective question, we condensed the answer scale (0-10) into four categories as well. The actigraphic variable sleep efficiency was categorized based on deciles, and we condensed the fourth subjective question scale (0-10) into ten categories. Histograms showing the distributions of the responses for the four actigraphic sleep variables are shown in Figure 5.

2.3 Statistical Analyses

2.3.1 Descriptive Statistics

Descriptive statistical analysis was done using Stata 15 (StataCorp, College Station, TX) to observe the characteristics of the dataset (Table 1), to understand the sleep patterns of the participants, and to examine possible differences between the questionnaire and actigraphic data (Table 2). All continuous variables were examined using mean and standard deviation. Categorical variables were tabulated and displayed as number of subjects in each category with the respective percentage.

2.3.2 Testing Concordance

We used a three-level random intercept model as the primary measurement of concordance, with level-one as the timepoint at which sleep was measured, level-two as the method of sleep measurement, and level-three as the subjects of the study (Figure 2). For these analyses, the main variables used in each model included subject ID, timepoint, a subjective question and the corresponding actigraphic sleep measurement. When the data are in long format, an additional variable for the combination of sleep measurement methods is included. This model allows us to assess the concordance of the two methods while controlling for the between-method within-subject heterogeneity by including both a random intercept for each combination of method and subject and a random intercept for subject. The model is given by:

$$y_{ijk} = \beta + \zeta_{jk}^{(2)} + \zeta_k^{(3)} + \varepsilon_{ijk}$$

In this model y_{ijk} is defined as the outcome of subject k for timepoint i for method j , β is the overall mean of y_{ijk} , $\zeta_{jk}^{(2)}$ is the random intercept for method j and subject k , $\zeta_k^{(3)}$ is the random intercept for subject k , and ε_{ijk} is the random error term. The sources of variability in this model include: (1) level-one variance, θ , which is the between-timepoint, within-method, and within-subject variance, (2) level-two variance, $\psi^{(2)}$, which is the between-methods, within-subject variance, and (3) level-three variance, $\psi^{(3)}$, which is the between-subject variance. Residual diagnostics were done after the models were fit using Empirical Bayes prediction for random intercepts at the subject level, random intercepts at the method level, and level-one residuals in the form of histograms and quantile-quantile plots. For each model to be considered a good fit, the quantile-quantile-plots should follow a straight line, and the histograms should follow a normal distribution.

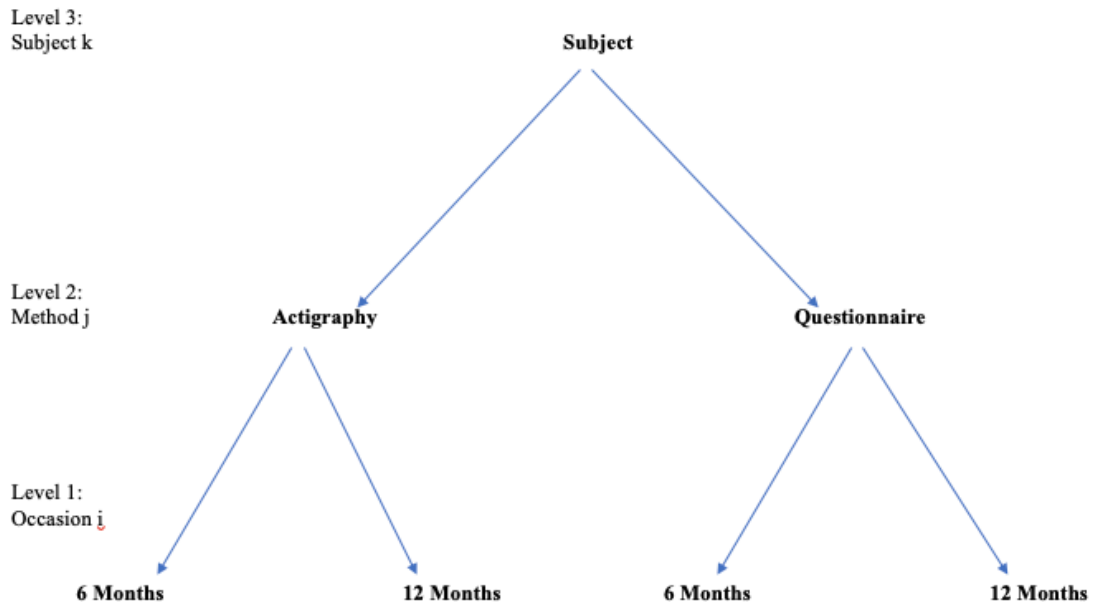


Figure 2. Three-level random intercept model.

After fitting this model for each of the different sleep measurements, the intraclass correlations (ICCs) can be calculated. In a three-level model, the types of ICCs include: (1) an ICC for both timepoints and the same subject using different measurement methods, (2) an ICC for both timepoints, on the same subject, and the same measurement method, and (3) an ICC for subject relative to measurement method (ignoring timepoint). The level-1 ICC calculation takes into account different timepoints and different methods of sleep measurement for the same subject. A high level-1 ICC implies the most variation is due to timepoint because both method and timepoint vary, which means that most variation is attributed to the lowest level. The level-2 ICC calculation takes into account different timepoints for the same subject and the same method. A high level-2 ICC implies the variation is due to the method of sleep measurement being used because variation in time point has already been considered. The level-3 ICC calculation takes into account different methods of sleep measurement for same subject while ignoring the timepoint variable. A high level-3 ICC implies the most variation is due to subject because utilizing different methods has already been considered. All of these types of ICCs were calculated and assessed for each of the models that were fit. The types of ICCs can be estimated by:

$$\begin{aligned}
 1) \quad \rho(\text{subject}) &= \frac{\psi^{(3)}}{\psi^{(2)} + \psi^{(3)} + \theta} \\
 2) \quad \rho(\text{method, subject}) &= \frac{\psi^{(2)} + \psi^{(3)}}{\psi^{(2)} + \psi^{(3)} + \theta} \\
 3) \quad \rho(\text{method}) &= \frac{\psi^{(3)}}{\psi^{(2)} + \psi^{(3)}}
 \end{aligned}$$

In order to perform inference on the variance components of the three-level random intercept models, two multilevel mixed effects maximum likelihood models are fit, one with the variable for the combination of measurement methods and one without this variable, and then a

likelihood ratio test is conducted to test if the between-method within-subject variance is significantly different from zero at a significance level set at $p < 0.05$.

Bland-Altman plots were used as a visual method to assess concordance. This plot, also known as a difference plot, is commonly used to compare different measurement techniques. The graph plots the difference between the measurements versus the averages of the two methods. The horizontal lines on the graph denote the mean difference and the limits of agreement, which is the 95% confidence interval for the mean difference.¹⁶ A Bland-Altman plot was created for each comparison of sleep measurement variables and assessed for concordance with bar charts included for the categorical variables that have a limited range.

2.3.3 Cross-lag Regression

To understand if changes in the subjective data preceded changes in the actigraphy data, we used cross-lag regression analysis using structural equation models. Cross-lag regression allows us to evaluate the directional effect variables have on each other over a period of time in order to establish which variable is causally predominant, while controlling for correlations within timepoints and autoregressive effects. This model assumes synchronicity, stationarity, variables are measured without error, X_1 occurs before X_2 , no confounders exist, standardized variables, and stability.¹²

Cross-lag models compare the relationship of the independent variable, X , at the first timepoint (X_1) and the dependent variable, Y , at the second timepoint (Y_2) to Y at the first timepoint (Y_1) and X at the second timepoint (X_2).¹² For this thesis, X_1 will represent the subjective measurement at six months, X_2 will represent the subjective measurement at twelve months, Y_1 will represent the objective measurement at six months, and Y_2 will represent the

objective measurement at twelve months. A cross-lag regression model was run for each comparison of sleep measurement variables and assessed for causal predominance. After these analyses were performed, goodness of fit was assessed for all four models. The three main goodness of fit measures we assessed were the Tucker Lewis Index (TLI), the comparative fit index (CFI), and the root mean squared error of approximation (RMSEA).

3.0 Results

3.1 Descriptive Statistics

The sample for this current study consisted of 137 participants, who were predominantly female (89.8%) and white (81.6%). The mean age was 51.8 years, with a standard deviation of 9.75. When asked how much trouble a person had falling asleep, the mean answer was 1.52 with a standard deviation of 0.84, which implies that most people had a relatively easy time falling asleep. For the number of minutes it took someone to fall asleep based on the actigraphic watch data, the mean was 7.62 with a standard deviation of 12.94. The mean number of minutes of sleep reported in the questionnaire was 418.1 with a standard deviation of 74.41, and the mean number of minutes of sleep captured by the actigraphic watch was 411.04 with a standard deviation of 80.19. The mean number of times a person awoke based on the questionnaire was 1.62 with a standard deviation of 1.5, and the mean number of wake bouts based on the actigraphic watch was 32.97 with a standard deviation of 14.21. When asked how well participants had slept (0 = poor, 10 = excellent), the mean answer was 6.41 with a standard deviation of 2.02. The mean sleep efficiency measure for the actigraphic watch was 88.95 with a standard deviation of 5.73. Frequencies were reported for the categorical variables. Summary statistics of the variables can be found in Table 1, and the number of participants in each category of the categorized subjective and objective variables can be found in Table 2.

Table 1. Descriptive statistics for demographic and sleep variables.

<u>Variable</u>	<u>n (%)</u>
<u>Time Point</u>	
6 months	1563 (49.24)
12 months	1611 (50.76)
<u>Sex</u>	
Male	320 (10.08)
Female	2854 (89.92)
<u>Race</u>	
White	2591 (81.63)
Non-white	583 (18.37)
	Mean (sd)
Age	51.8 (9.75)
Subjective Question 1: Did you have trouble falling asleep? (0 = no trouble, 10 = a lot of trouble)	1.52 (0.84)
Subjective Question 2: How many hours of sleep did you get? (minutes)	418.1 (74.41)
Subjective Question 3: Number of awakenings?	1.62 (1.50)
Subjective Question 4: Rate how well you slept last night. (0 = poor, 10 = excellent)	6.41 (2.02)
Objective (Actigraphy) 1: Sleep onset latency	7.62 (12.94)
Objective (Actigraphy) 2: Sleep time in minutes	411.04 (80.19)
Objective (Actigraphy) 3: Number of wake bouts	32.97 (14.21)
Objective (Actigraphy) 4: Sleep efficiency	88.95 (5.73)

Table 2. Categorized variable counts.

Variable	n (%)	Variable	n (%)
<u>Subjective Question 1:</u> <u>Did you have trouble falling asleep?</u> (1 = no trouble, 4 = a lot of trouble)		<u>Objective Actigraphy 1:</u> <u>Sleep onset latency</u>	
1 [0, 3]	1717 (54.10)	1 [0, 0.5]	984 (31.00)
2 [4, 5]	386 (12.16)	2 (0.5, 3]	610 (19.22)
3 [6, 8]	363 (11.44)	3 (3, 9.5]	817 (25.74)
4 [9, 10]	72 (2.27)	4 (9.5, 170]	763 (24.04)
NA	636 (20.04)		
<u>Subjective Question 4:</u> <u>Rate how well you slept last night</u> (1 = poor, 10 = excellent)		<u>Objective Actigraphy 4:</u> <u>Sleep efficiency</u>	
1 [0,1]	70 (2.21)	1 [36.2, 82]	318 (10.02)
2 [2]	73 (2.30)	2 (82, 85.6]	317 (9.99)
3 [3]	110 (3.47)	3 (85.6, 87.6]	317 (9.99)
4 [4]	153 (4.82)	4 (87.6, 89]	318 (10.02)
5 [5]	288 (9.07)	5 (89, 90.3]	317 (9.99)
6 [6]	422 (13.30)	6 (90.3, 91.2]	318 (10.02)
7 [7]	621 (19.57)	7 (91.2, 92.2]	317 (9.99)
8 [8]	482 (15.19)	8 (92.2, 93.3]	318 (10.02)
9 [9]	233 (7.34)	9 (93.3, 94.5]	317 (9.99)
10 [10]	79 (2.49)	10 (94.5, 99.3]	317 (9.99)
NA	643 (20.26)		

3.2 Testing Concordance

Four three-level random intercept models were analyzed, one for each subjective question and its actigraphic counterpart. The intraclass correlations (ICCs) are summarized in Table 3. The level-1 ICCs, $\rho(\text{subject})$, explain the percentage of the total variation in the dependent variable that is attributable to the variation between timepoints. The level-2 ICCs, $\rho(\text{method}, \text{subject})$, explain the percentage of the total variation in the dependent variable that is attributable to the variation

between methods. The level-3 ICCs, $\rho(\text{method})$, explain the percentage of the total variation in the dependent variable that is attributable to the variation between subjects. For the first model, the factor that appears to have the largest influence on the total variation is the method that is used. For the second model, the factor that appears to have the largest influence on the total variation is the subject. For the third and fourth models, the factor that appears to have the largest influence on the total variation is also the method that is used.

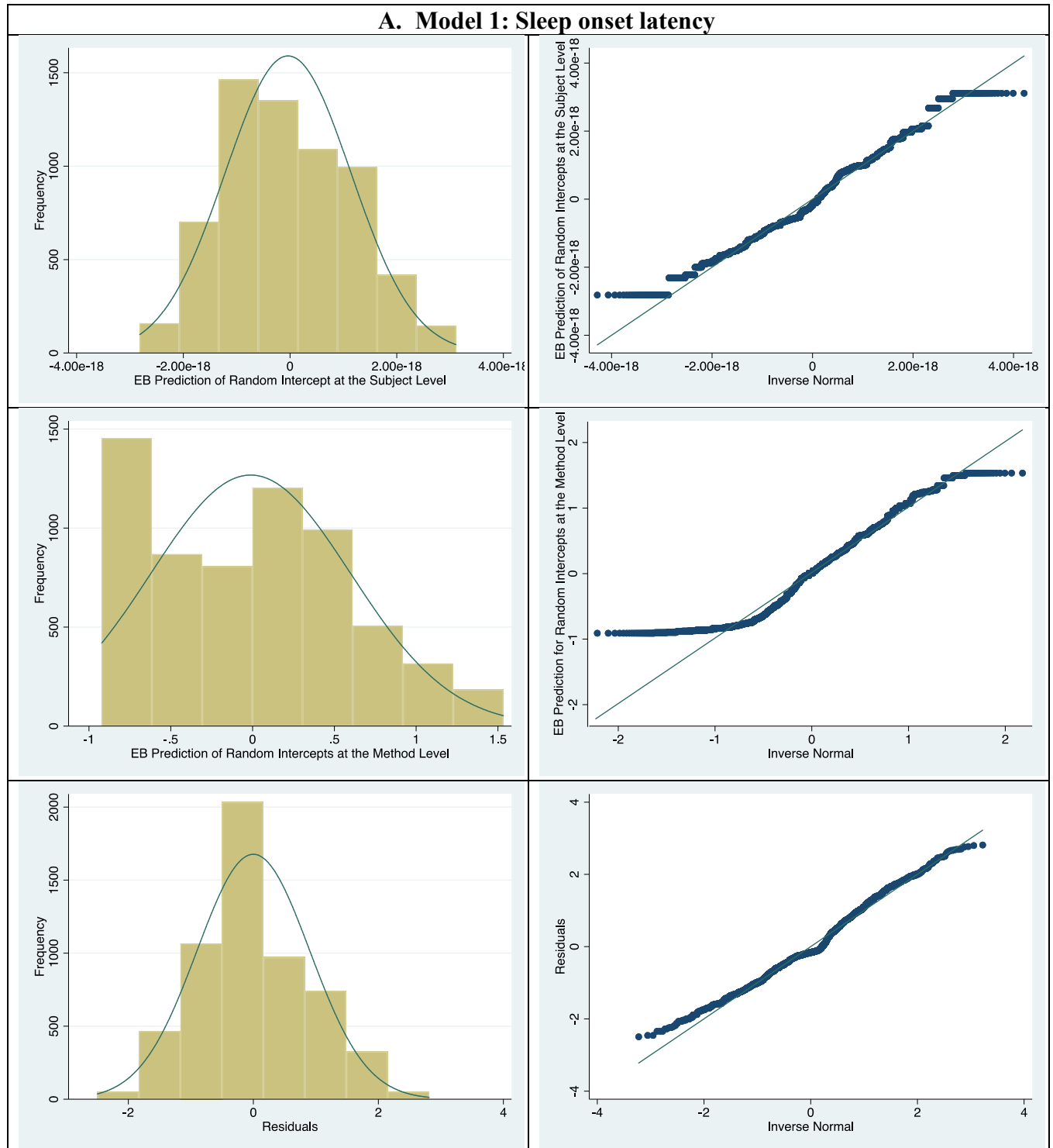
Table 3. Intraclass correlations for each model.

Intraclass Correlation (ICC)	Model 1: Sleep onset latency	Model 2: Sleep time	Model 3: Number of wake bouts	Model 4: Sleep efficiency
$\rho(\text{subject})$	<0.0001	0.2176	<0.0001	0.0508
$\rho(\text{method,subject})$	0.3264	0.3191	0.8277	0.3973
$\rho(\text{method})$	<0.0001	0.6819	<0.0001	0.1278

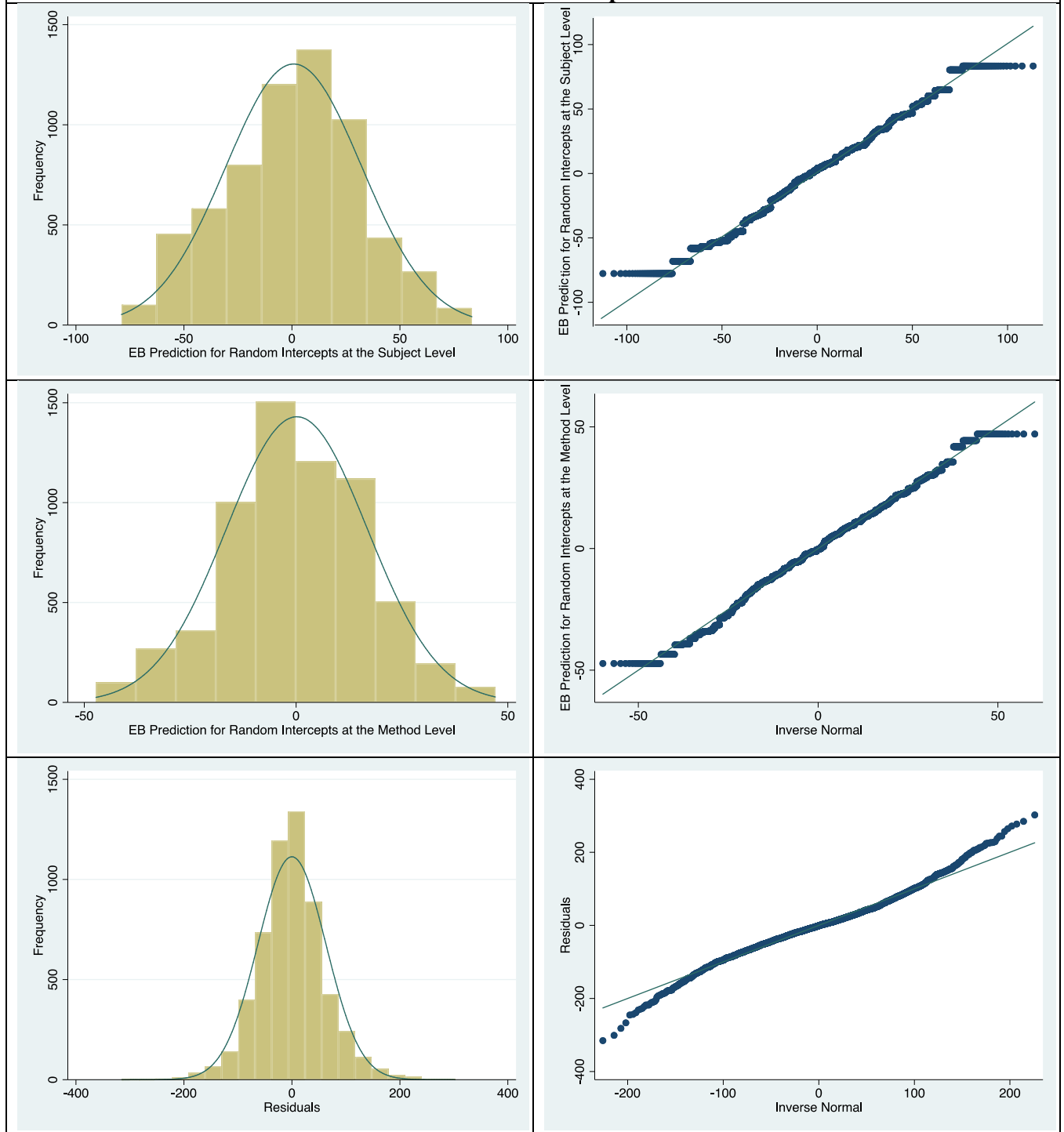
Two different nested three-level models were analyzed both with and without the method of sleep measurement as a factor. The likelihood ratio test used to compare the models at a significance level of $p < 0.05$ showed that the between-method within-subject variance is statistically significant different from zero for each of the four models that were tested. This suggests that the objective and subjective methods of sleep measurement are not concordant for any of the four sleep measurement models tested because the variance between the methods is greater than zero. Empirical Bayes predictions for random intercepts at the subject level, random intercepts at the method level, and level-one residuals in the form of histograms and quantile-quantile plots can be seen below in Figure 3. Based on the plots for model one, the residuals look relatively normal, which indicates a good model fit. The random intercepts at the method level look skewed in the histogram, but the quantile-quantile plot shows only a slight deviation. All plots look normal in model two, indicating a good model fit. For model three, there are a few

outliers at the subject level, but the residual plots are mostly normal. At the method level, the data looks heavily skewed, and the level-one residuals look normal, indicating a good model fit.

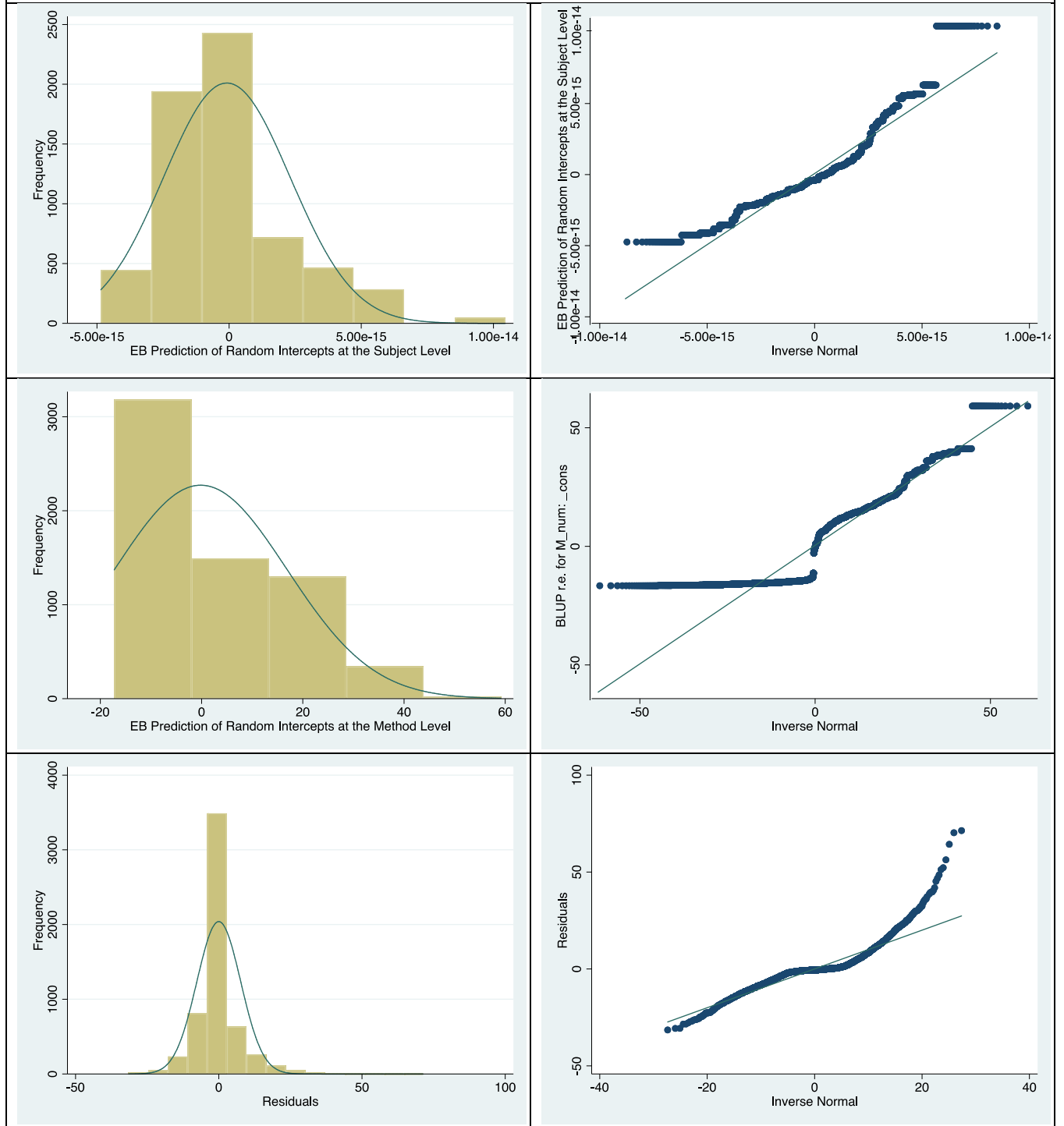
All plots look normal in model four, indicating a good model fit.



B. Model 2: Sleep time



Model 3: Number of wake bouts



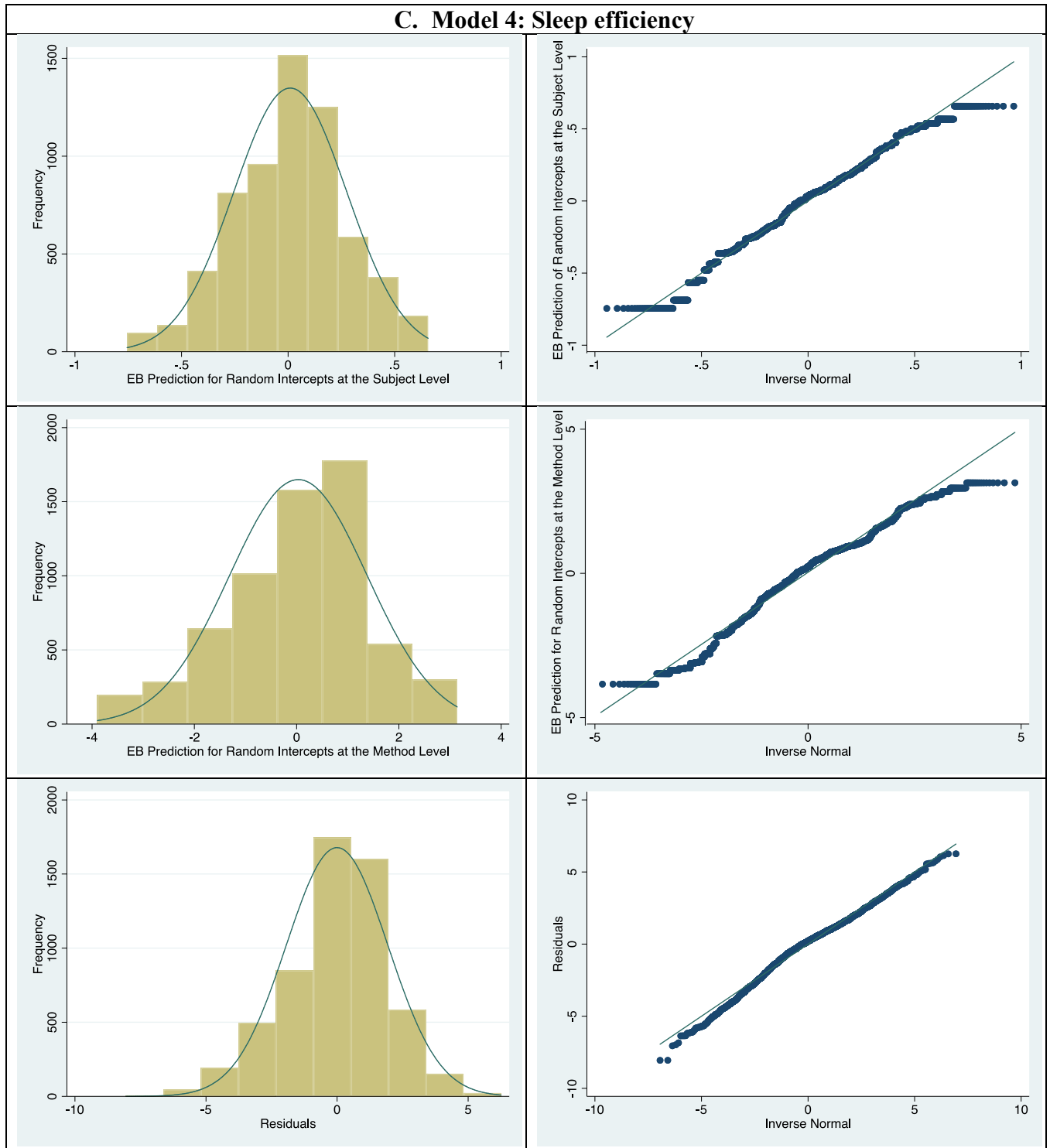


Figure 3. Residual diagnostics of three-level random intercept model using Empirical Bayes prediction.

Bland-Altman plots were created for each of the four models (Figure 4). Plots 4A and 4D have a similar pattern, which is because these two models that include categorized data, and the limits of agreement are wide for the categorized variables. In order to supplement the information given by the Bland-Altman plot, bar graphs were also included. The bar plots show that the difference between the two methods of sleep measurement is only zero about 30.54% of the time for sleep onset latency, and the difference between the two methods of sleep measurement is only zero about 13.24% of the time for sleep efficiency. The mean difference between the two methods is greater than zero for all of the points for the number of awakenings measurements. While the amount of time asleep does have a large clustering within the limits of agreement, there is still a large number of points outside of these limits. All four of these plots appear to agree with the conclusion that the two methods of sleep measurement are not concordant for any of the four sleep measurement models that were tested.

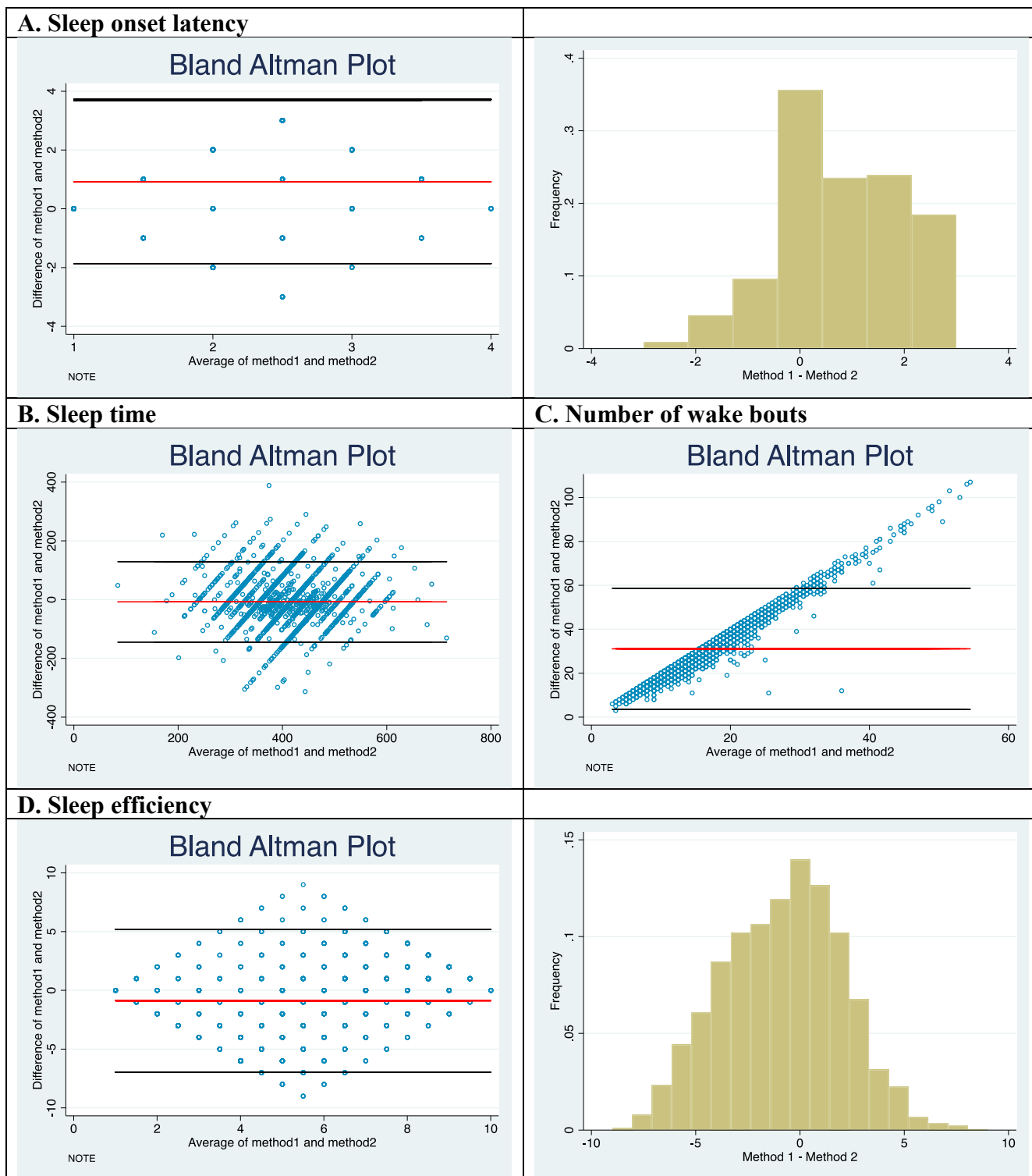


Figure 4. Bland-Altman plots that show the average between methods versus the difference between methods.

3.3 Cross-lag Regression

The coefficients (β) and p-values from each of the cross-lag regression models along with their corresponding goodness of fit tests are summarized in Table 4. For model one, the covariance for the objective measurements and the subjective measurements is about -0.003, which shows a very small, negative relationship. For model two, the covariance is about -142.24, which shows a large, negative relationship. For model three, the covariance is about 1.22, which shows a small, positive relationship. For model four, the covariance is about -0.24, which shows a small negative relationship.

The p-values of the coefficients indicate whether or not there is a statistically significant relationship occurring at a significance level of $p < 0.05$. For model three, β_1 is statistically significant, which indicates that the subjective measurement at the timepoint of six months has a statistically significant relationship to the objective measurement at the timepoint of twelve months for the participants' number of awakenings. Similarly for model four, β_1 is statistically significant, which indicates that the subjective measurement at the timepoint of six months has a statistically significant relationship to the objective measurement at the timepoint of twelve months for the participants' quality of sleep. These two findings suggest that some changes in subjective sleep precede changes in objective sleep for the number of times participants wake after falling asleep and for their overall quality of sleep. The goodness of fit measurements all show that the model is a good fit because the RMSEA should be less than 0.05, and the CFI and TLI should be close to one, however, the results being exactly zero, one, and one respectively for these goodness of fit measurements may be due to having a saturated model, which means the data will fit perfectly at any point and goodness of fit measurements are not necessary. A saturated model is not a concern

if the model fits the theoretical perspective of the analysis. A fully-saturated model is the only way to obtain both autoregressive and cross-lagged coefficients for a cross-lagged regression model with two variables at two timepoints. Both types of coefficients are necessary to fully assess causal predominance.¹⁷

Table 4. Summary of cross-lag regression for each model.

Model Parameters	Model 1: Sleep onset latency	Model 2: Sleep time	Model 3: Number of wake bouts	Model 4: Sleep efficiency
β_1(p-value)	0.04(0.218)	-0.02(0.523)	0.09(0.002)	0.09(0.002)
β_2(p-value)	0.01(0.679)	0.03(0.203)	0.03(0.314)	0.02(0.445)
β_3(p-value)	0.02(0.347)	0.05(0.053)	-0.005(0.110)	-0.01(0.659)
β_4(p-value)	0.05(0.224)	-0.05(0.100)	-0.47(0.103)	0.05(0.219)
Covariance (p-value)	-0.003(0.924)	-142.24(0.405)	1.22(0.071)	-0.24(0.156)
RMSEA	0	0	0	0
CFI	1	1	1	1
TLI	1	1	1	1

4.0 Discussion

This analysis suggests that objective and subjective sleep measurement techniques are not concordant. This conclusion agrees with several other similar studies. Prior studies have found that there is a higher concordance between the objective measures of actigraphy and PSG than between either objective method and the subjective method of sleep diaries.³ Another study found that electronic diaries show a significantly longer sleep duration than actigraphy, and the self-reported sleep quality had little to no relationship with actigraphy.¹⁸ Several other studies show that actigraphy tends to underestimate sleep when compared to subjective measurements of sleep, while others claim that actigraphy overestimates sleep, but either way, they are discordant.^{4, 19}

This thesis utilized subjective sleep questions, which we were able to match with actigraphic sleep measurements in order to test concordance. While there was a match for each question with an actigraphic variable, categorization of some of the variables was necessary in order to make this analysis interpretable. In the future, it may be beneficial to compare objective and subjective methods of sleep on a more similar scale to enable a more accurate picture of concordance. For instance, the measurements were very similar in this thesis for model two and model three because the data were both continuous. The subjective questions that were answered on a scale of 0-10 required us to categorize the continuous actigraphic variables. An additional question, ‘Do you feel tired? (y/n)’ was also answered by participants during data collection, but this binary variable would have required us to divide a continuous actigraphic variable in half for interpretability, which would lead to inconclusive results. In the future, the subjective questions could be rephrased in order to obtain answers that are continuous, which may lead to a more precise analysis.

The cross-lag regression methods used here were helpful in understanding if one method of measurement at one timepoint influences another. This should not, however, be interpreted as a causal relationship. It has been suggested that cross-lag regression results should be interpreted strictly in terms of influence and not cause and effect.¹² This method is often used to try to understand predictive relationships, but the model itself does not take the time interval between measurements into account, which can lead to biased results when the time intervals being observed are unequal. If the time interval between measurements is constant, the study results are only applicable to an interval of that size.²⁰ In this analysis, we did find two statistically significant relationships, but it is important to emphasize that this is an influential relationship that is only applicable to a six month period of time.

In future analysis, a more diverse study sample would strengthen results. The sample used in this thesis was predominately female and white. The age group in this sample was mostly middle-aged adults, so it could be interesting to repeat this analysis with a younger population. A study on an adolescent sample found that there was higher discordance between actigraphy and self-reported sleep among the boys, which showed a discordance of about two hours, whereas girls showed a discordance between the two methods of about thirty minutes.²¹

Appendix: Statistical Code

A.1 Additional Figures

A.1.1 Histograms of the Objective Sleep Measurement Variables

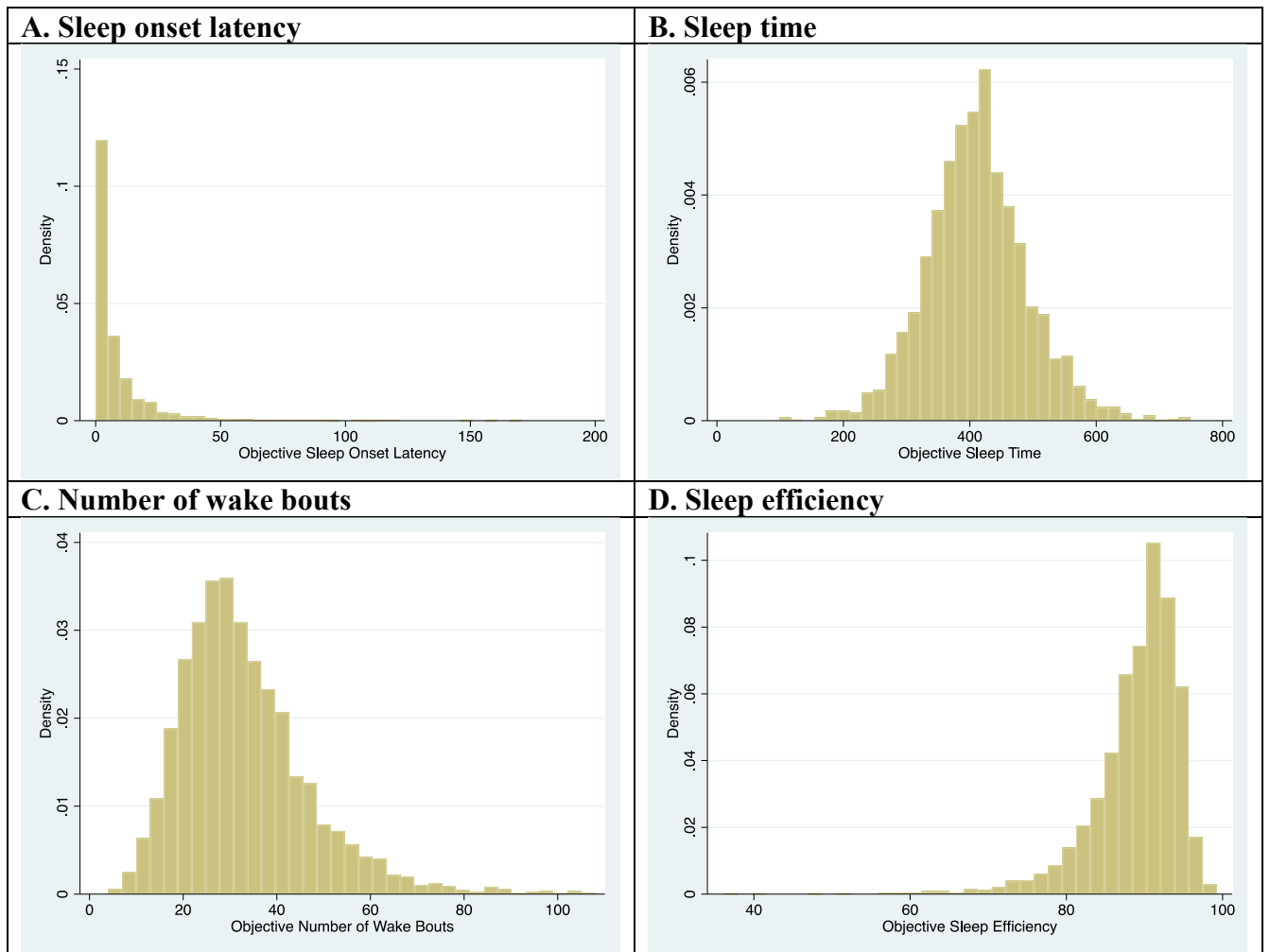


Figure 5. Distributions of the objective sleep measurements.

A.2 Stata Example Code

A.2.1 Stata example code for descriptive analysis

```
summarize(age Q46 Q47 Q48 Q49 sleep_onset_latency sleep_sleep_time  
sleep_number_of_wake_bouts sleep_efficiency)  
tabulate(timepoint sex race_2cat QSOL QSE stratSOL stratSE)
```

A.2.2 Stata example code for three-level random intercept model

```
gen i = _n  
gen method1 = sleep_sleep_time  
gen method2 = nQ47  
drop nQ47  
drop sleep_sleep_time  
reshape long method, i(i) j(M) string  
encode M, gen(M_num)  
mixed method || ID: || M_num:, mle stddev nolog  
predict rsubj rM, reffects  
predict yhat, fitted  
quietly mixed method || ID: || M:, mle  
estimates store threelevel  
quietly mixed method || ID:, mle  
lrtest threelevel  
predict resid, residuals  
hist rsubj, freq normal  
hist rM, freq normal  
hist res, freq normal  
qnorm rsubj  
qnorm rM  
qnorm res
```

A.2.3 Stata example code for Bland-Altman plots²²

```
set more off  
capture program drop blandaltman  
program blandaltman  
syntax varlist(max=2)  
  
// prepare for Bland Altman Interreader
```



```

tempvar diff_xy
tempvar avg_xy
tempvar lower
tempvar higher
tempvar MW
tempvar SE
tempvar CIhigher
tempvar CIlower

generate `diff_xy'=0
generate `avg_xy'=0
generate `lower'=0
generate `higher'=0
generate `MW'=0
generate `SE'=0
generate `CIhigher'=0
generate `CIlower'=0

// count the variable: how many variable are in the list?
local noofvars : word count `varlist'
display as text "The variable list of this program counts "
`noofvars' " variables"
display as result " "
display as result " "

// Interreader
local x = 1
local y = 1
foreach varx of varlist `varlist' {
  foreach vary of varlist `varlist'{
    if `y' > `x'{
      quietly replace `avg_xy'=(`varx'+`vary')/2
      quietly replace `diff_xy'=`varx'-`vary'
      display as result " Bland Altman Plot of `varx' and `vary'"
      quietly sum `diff_xy'
      quietly return list
      quietly replace `MW'=r(mean)
      quietly replace `lower'=r(mean)-2*r(sd)
      quietly replace `higher'=r(mean)+2*r(sd)
      quietly replace `SE'=(r(sd))/(sqrt(r(N)))
      quietly replace `CIlower'=r(mean)-2*`SE'
      quietly replace `CIhigher'=r(mean)+2*`SE'
      display as result "- mean of difference between `varx' and `vary'
      is "r(mean)
      display as result "- sd of difference between `varx' and `vary' is
      "r(sd)
    }
    ++y
  }
  ++x
}

```

```

display as result "- lower limit of difference between `varx' and
`vary' is " `lower'
display as result "- higher limit of difference between `varx' and
`vary' is " `higher'
display as result "- Limits of agreement (Reference Range for
difference): " `lower' " to " `higher'
display as result "- Mean difference:" `MW' " (CI " `CIlower' " to
" `CIhigher' ")
display as result " "
display as result " "

label var `diff_xy' "Values"
label var `MW' "mean of difference"
label var `lower' "lower limit of agreement"
label var `higher' "higher limit of agreement"
twoway (scatter `diff_xy' `avg_xy', msymbol(smcircle_hollow)
mcolor(ebblue)) (line `MW' `avg_xy', lcolor(red)) (line `lower'
`avg_xy', lcolor(black) ) (line `higher' `avg_xy', lcolor(black)
), title(Bland Altman Plot, size(8)) subtitle(,size(5))
xtitle(Average of `varx' and `vary') ytitle(Difference of `varx'
and `vary') caption() note(NOTE) legend(off)
}
local y = `y'+1
}
local y = 1
local x = `x'+1
}
end
gen method1 = sleep_sleep_time
gen method2 = nQ47
drop sleep_sleep_time
drop nQ47
blandaltman method1 method2

```

A.2.4 Stata example code for cross-lag regression analysis

```

destring timepoint, generate(time) ignore(`"MO"', illegal)
gen method1 = sleep_sleep_tie
gen method1 = sleep_sleep_time
gen method2 = nQ47
gen objtime1 = method1 if time == 6
gen objtime2 = method1 if time == 12
gen subjtime1 = method2 if time == 6
gen subjtime2 = method2 if time == 12

```

```
import                                                    delimited
/Users/jocelynmineo/Documents/CrossLagModel1Fixed.csv,
case(preserve) clear
sem (objtime2 <- objtime1 subjtime1) (subjtime2 <- objtime1
subjtime1), cov(e.objtime2*e.subjtime2)
estat gof, stats(all)
```

Bibliography

1. Naitoh, Paul, et al. *Health Effects of Sleep Deprivation*. Hanley & Belfus, Inc., 1990, apps.dtic.mil/dtic/tr/fulltext/u2/a249653.pdf.
2. Taibi DM; Landis CA; Vitiello MV. Concordance of polysomnographic and actigraphic measurement of sleep and wake in older women with insomnia. *J Clin Sleep Med* 2013;9(3):217-225.
3. Mundt JM, Crew EC, Krietsch K, Roth AJ, Vathauer K, Robinson ME, Staud R, Berry RB, McCrae CS. Measuring treatment outcomes in comorbid insomnia and fibromyalgia: concordance of subjective and objective assessments. *J Clin Sleep Med* 2016;12(2):215–223.
4. Kushida, Clete A, et al. “Comparison of Actigraphic, Polysomnographic, and Subjective Assessment of Sleep Parameters in Sleep-Disordered Patients.” *Sleep Medicine*, vol. 2, no. 5, 21 Aug. 2000, pp. 389–396., doi:10.1016/s1389-9457(00)00098-8.
5. Shiffman, Saul. “Ecological Momentary Assessment.” *Oxford Handbooks Online*, 2014, doi:10.1093/oxfordhb/9780199381708.013.1.
6. Zheng, Y, et al. “Patterns of Self-Weighing Behavior and Weight Change in a Weight Loss Trial.” *International Journal of Obesity*, vol. 40, no. 9, 2016, pp. 1392–1396., doi:10.1038/ijo.2016.68.
7. Kline, Christopher E., et al. “Bidirectional Relationships Between Weight Change and Sleep Apnea in a Behavioral Weight Loss Intervention.” *Mayo Clinic Proceedings*, vol. 93, no. 9, Sept. 2018, pp. 1290–1298., doi:10.1016/j.mayocp.2018.04.026.
8. Sereika, Susan M., et al. “Modern Methods for Modeling Change in Obesity Research in Nursing.” *Western Journal of Nursing Research*, vol. 39, no. 8, 2017, pp. 1028–1044., doi:10.1177/0193945917697221.
9. Mendez, Dara D., et al. “Neighborhood Factors and Six-Month Weight Change among Overweight Individuals in a Weight Loss Intervention.” *Preventive Medicine Reports*, vol. 4, 20 Oct. 2016, pp. 569–573., doi:10.1016/j.pmedr.2016.10.004.
10. Snijders, Tom A.b., and Johannes Berkhof. “Diagnostic Checks for Multilevel Models.” *Handbook of Multilevel Analysis*, 2007, pp. 141–175., doi:10.1007/978-0-387-73186-5_3.
11. Smith, Mark W., et al. “Bar Charts Enhance Bland–Altman Plots When Value Ranges Are Limited.” *Journal of Clinical Epidemiology*, vol. 63, no. 2, Feb. 2010, pp. 180–184., doi:10.1016/j.jclinepi.2009.06.001.

12. Kearney, Michael. (2017). Cross-Lagged Panel Analysis.
13. Kenny, David. (1975). Cross-Lagged Panel Correlation: A Test for Spuriousness. *Psychological Bulletin*. 82. 887-903. 10.1037/0033-2909.82.6.887.
14. Allison, Paul. (2017). Longitudinal Data Analysis Using Structural Equation Modeling. Statistical Horizons Seminar.
15. Burke, Lora E, et al. "Ecological Momentary Assessment in Behavioral Research: Addressing Technological and Human Participant Challenges." *Journal of Medical Internet Research*, vol. 19, no. 3, 15 Mar. 2017, doi:10.2196/jmir.7138.
16. Schoonjans, Frank. "Bland-Altman Plot." *MedCalc*, MedCalc Software, 18 Jan. 2019, www.medcalc.org/manual/blandaltman.php.
17. Labhart, Florian, et al. "Reciprocal Influences of Drinking Motives on Alcohol Use and Related Consequences: A Full Cross-Lagged Panel Study Among Young Adult Men." *Behavioral Medicine*, vol. 43, no. 4, 9 Mar. 2016, pp. 277–284., doi:10.1080/08964289.2016.1157057.
18. Okifuji, Akiko, and Bradford D. Hare. "Nightly Analyses of Subjective and Objective (Actigraphy) Measures of Sleep in Fibromyalgia Syndrome." *The Clinical Journal of Pain*, vol. 27, no. 4, May 2011, pp. 289–296., doi:10.1097/ajp.0b013e31820485db.
19. Short, Michelle A., et al. "The Discrepancy between Actigraphic and Sleep Diary Measures of Sleep in Adolescents." *Sleep Medicine*, vol. 13, no. 4, Apr. 2012, pp. 378–384., doi:10.1016/j.sleep.2011.11.005.
20. Kuiper, Rebecca M., and Oisín Ryan. "Drawing Conclusions from Cross-Lagged Relationships: Re-Considering the Role of the Time-Interval." *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 25, no. 5, 5 Mar. 2018, pp. 809–823., doi:10.1080/10705511.2018.1431046.
21. Guedes, Luciane Gaspar, et al. "Comparison between Self-Reported Sleep Duration and Actigraphy among Adolescents: Gender Differences." *Revista Brasileira De Epidemiologia*, vol. 19, no. 2, June 2016, pp. 339–347., doi:10.1590/1980-5497201600020011.
22. Caviezel, Seraina. "Bland Altman Plot." *ResearchGate*, ResearchGate, 11 Nov. 2013, www.researchgate.net